DOCUMENT RESUME

ED 353 264 TM 018 905

AUTHOR Hodges, Carol A.
TITLE Teacher Judgments and Standardized Assessments.
PUB DATE Apr 92
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Comparative Analysis; Elementary School Students; *Elementary School Teachers; Evaluators; Grade 1; Grade 2; *Kindergarten Children; Literacy; Longitudinal Studies; Primary Education; Scores; *Standardized Tests; *Student Evaluation; *Teacher Role; Test Results; Test Use

ABSTRACT

This study, part of a larger longitudinal research project, focused on the relationship between teachers' rankings of their students' literacy achievements, based on informal assessment, and the scores that the students received on a standardized test. Kindergarten students (initial sample of 136) were followed for 3 years. At the end of kindergarten, grade 1, and grade 2, teachers evaluated the students' progress according to how well they mastered a set of criteria that represented a successful reader and writer. Teachers administered a standardized test to students each year. Data from all 3 years illustrated significant positive relationships between teachers' evaluations and test scores. The fact that the teachers and the tests appear to be measuring several similar factors should ease fears that teacher judgments might be at odds with that which is currently considered more reliable, test scores. However, data do suggest that teacher judgments may be an even more valid measure than standardized test scores. One table presents study data, 2 figures present the assessment forms for reading and writing, and 13 references are included. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED353264

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CAROL A. HODGES

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TEACHER JUDGMENTS AND STANDARDIZED ASSESSMENTS

paper presented at the

1992 AERA ANNUAL MEETING

by

Dr. Carol A. Hodges
Associate Professor
Buffalo State College

1992 AERA ANNUAL MEETING

TEACHER JUDGMENTS AND STANDARDIZED ASSESSMENTS

TM018905

BEST COPY AVAILABLE

ABSTRACT

The belief that data from standardized tests are more trustworthy than data collected by other means is often held by school administrators despite the fact that reading researchers have found that reading assessment has not kept pace with advances in research, theory, and practice and that early childhood experts agree that extensive standardized achievement testing narrows and misdirects the curriculum and drains instructional time.

This study, part of a larger longitudinal research project, focused on the relationship between teachers' rankings of their students literacy achievements, based on informal assessment and the scores that the students received on a standardized test. Kindergarten students were followed for three years. At the end of their kindergarten, first grade, and second grade school years, teachers evaluated the students' progress according to how well they mastered a set of criteria which represented a successful reader and writer at the end of each grade Then the teachers administered a standardized test to the students.

Data from all three years illustrated there were significant positive relationships between the teachers' evaluations and the test scores. Correlations ranged from .57-.90 ($p<.01$) The fact that the teachers and the test appear to be measuring a number of similar factors should ease fears that teacher judgments might be totally at odds with what currently are considered to be the more reliable, the test scores.

At the same time, knowing what we do about the negative factors associated with standardized testing in the primary grades and the fact that little use seems to be made of the test results, the data suggest that teacher judgments, based on knowledge of their students' development and knowledge of the processes involved in reading and writing, may be even more valid means of obtaining assessment information.

INTRODUCTION

The use of standardized tests has increased dramatically over the past few decades and the trend toward more testing seems likely to continue. However, as the emphasis on standardized tests has escalated, so have objections to them. A number of reading researchers (Edelsky & Harmon, 1988; Garcia & Pearson, 1991; Hodges, 1989, 1991; Squires, 1987; Teale, 1988, Valencia & Pearson, 1986) have pointed out that early reading assessment has not kept pace with advances in reading research, theory, and practice. At the same time early childhood experts (Bredekamp, 1986; Fairtest & NYPIRG, 1990; International Reading Association, 1986; Harmon, 1990; Moyer, Egertson, & Isenberg, 1987; National Association for the Education of Young Children, 1988) argue that children are being tested too early. They claim that young children are not good test takers; that the unfamiliar format leads to stress; that test results are influenced by the children's ability to sit still and be quiet; and that extensive testing narrows and misdirects the curriculum and drains instructional time without a clear demonstration that the investment is beneficial. In addition, groups as diverse as the American Association of Colleges for Teacher Education (AACTE), the American Federation of Teachers (AFT), the National Association of Elementary School Principals (NAESP), and the national PTA have spoken out to urge states to abandon the use of multiple-choice tests and to replace them with alternative assessment techniques which seek to measure directly the student's ability to perform in the subject area. Several years ago when I examined a variety of standardized readiness and early reading tests, I found that such tests:

- **have stayed the same for virtually sixty years**. They reflect behavioral research that assumes that literacy can only be taught through the direct instruction of isolated skills which are **hierarchically organized and mastered one level at a time**.

- are inappropriate for literacy assessment based on a holistic philosophy because they **measure only reading skills and ignore other components of literacy** such as speaking, listening, and writing, components which we now realize are highly correlated with success in reading. And even in the area of reading these tests measure only a **narrow range** of the knowledge and skills involved. For instance the readiness and early reading tests I surveyed focused on phonic skills and vocabulary recognition, ignoring other important aspects such as background knowledge, directionality, print awareness, other word recognition strategies, comprehension strategies, appreciation of books, and interest in reading and writing.

- **are divorced from the reality of authentic reading and writing situations**. They depend on multiple choice questions and the materials read consist of single letters, words, and sentences, or very brief passages. Furthermore, pupils background knowledge and interests are ignored; the tests are timed; and pupils are not allowed to consult with one another. Obviously, such materials and settings are as far from authentic reading and writing situations as possible!

Unfortunately, despite their apparent potential, the acceptance of alternative assessment techniques has been constrained by concerns about their lack of efficiency and objectivity. Because it is easy to get a false sense of security when skilled reading is equated with scores on reading tests, many school personnel believe that data from standardized tests are more trustworthy than data collected by other means.

THE PROBLEM

After completing the survey of early reading tests, I began investigating the primary level (K-2 grade) literacy program and assessment tools of a school district in a small suburban community. This district tested all of its students beginning in kindergarten each May with a widely used standardized test battery. After interviewing the kindergarten teachers I discovered that they administered standardized reading achievement tests to students very reluctantly. They resented the time that the administration of the test took from instruction, the pressures that it put on the curriculum, and the frustration that it exerted on their students. In addition, because these teachers had made a transition from a basal readiness program to a more developmentally based process oriented literacy program, they felt the need to have a variety of assessment tools for the everyday instructional decision-making that is a crucial part of that approach. But they were not sure how to use informal assessment and, even they, wondered whether the informal tools could provide valid and reliable data.

The questions most often asked by the administration and the teachers were, "How would the teachers' assessment of students based on alternative evaluation techniques compare to the way in which the standardized test assessed them? How valid would teacher judgments be? The results of this study begin to provide information to answer those questions.

THE STUDY

Teacher Ratings: Year One

In May of 1989, before they administered the standardized achievement test, I asked seven kindergarten teachers to evaluate their 136 students according to how well they had mastered a set of criteria which the teachers felt represented the successful reader and writer at the end of kindergarten. Among the criteria reported by the teachers were the following:

- attitude toward books and reading/writing
- recognition of the letters of the alphabet
- knowledge of grapheme/phoneme correspondences
- use of invented spelling in writing
- ability to listen to and comprehend stories
- ability to read independently
- general maturity (following directions and keeping to a task)

The teachers assessed their students as being above average readers/writers (3), average readers/writers (2), and below average readers/writers (1) based on these criteria. They used a variety of assessment techniques including anecdotal records, observation checklists, and work samples. See Figures One and Two for examples of checklists used by the kindergarten teachers.

(Insert Figures One and Two about here)

The standardized test which the teachers later administered to their kindergartners purported to assess skills in auditory discrimination, grapheme/phoneme correspondence, decoding, and listening comprehension. The test scores examined were the Total Reading stanine scores (9-1) which the students earned.

A comparison of the teacher assessments with the Total Reading stanines reported on the standardized test showed that there was a significant relationship between the assessments of the students by the teachers and the Total Reading stanines obtained by the students on the standardized test. The degree of the relationship between the teacher ranked groups and the test scores was

computed by using the Pearson Product Moment correlation coefficient. Table one illustrates that the correlations for the classes ranged from .58-.87 ($p < .01$) A correlation of .71 ($p < .01$) was found over all classes.

(Insert Table One about here)

Teacher Ratings: Year Two

Near the end of these same students' first grade experience (May 1990), I asked their six first grade teachers to evaluate them according to how well they had mastered a set of criteria which the teachers felt represented the successful reader and writer at the end of first grade. Among the criteria reported by the teachers were the following:

- ability to handle books
- knowledge of how print works
- attitude toward books and reading/writing
- knowledge of grapheme/phoneme correspondences
- use of invented spelling in writing
- ability to listen to and comprehend stories
- ability to read independently

The teachers assessed their students as being above average readers/writers (3), average readers/writers (2), and below average readers/writers (1) based on these criteria. Over this second year the original population of 136 students was reduced to 117 because nineteen students who had been in the original kindergarten cohort were lost. They used a variety of informal assessment techniques.

The standardized test which the teachers later administered to their first graders was composed of subtests which assessed grapheme/phoneme correspondence, decoding, vocabulary, listening comprehension, and reading comprehension skills. Once more, the test scores examined were the Total Reading stanine scores (9-1) which the students earned.

A comparison of the teacher assessments of their students and the Total Reading stanines reported on the standardized test illustrated a significant relationship. (See Table One.) Correlations for the six classes ranged from .70-.84 ($p < .01$) A correlation of .67 ($p < .01$) was found over all teachers.

Teacher Ratings: Year Three

Near the end of the third year (1990-91), the students' six second grade teachers were also asked to assess their students reading/writing ability in the manner previously done by the kindergarten and first grade teachers. The set of criteria which the teachers felt represented the successful reader and writer at the end of second grade were the following:

- ability to handle books
- knowledge of how print works
- attitude toward books and reading/writing
- knowledge of grapheme/phoneme correspondences
- use of invented spelling in writing
- ability to listen to and comprehend stories
- ability to comprehend print in a variety of situations
- ability to read independently
- ability to follow directions

The subtests represented on the standardized test that year were letters and sounds, vocabulary, and comprehension. Seven more students from the original kindergarten cohort had left the school; leaving a total of 110 students whose progress was followed over the three years. They too used a variety of assessment techniques.

A comparison of the second grade teacher assessments with the Total Reading stanines reported on the standardized test again illustrated that there was a significant relationship between the assessments of the students by the teachers and the Total Reading stanines obtained by the students on the standardized test. (See Table One.) The degree of the relationship between the teacher assessed groups ranged from .63-.90 ($p < .01$)

Patterns of Differences: When the teacher assessments and the test scores are reviewed for individual teachers one interesting pattern becomes apparent. Because teachers were asked to evaluate their students as above average, average, or below average, they were in a sense encouraged to categorize some pupils in each class, as below average. The standardized test was not forced to do so. Therefore in some classes no children received a below average stanine score (1-3), but did receive a below average assessment by the teacher. Any replication of this study should word directions to teachers carefully so that they do not feel forced to place students in a below average category.

CONCLUSIONS AND IMPLICATIONS

What evidence would prove that teacher judgments can be valid measures of reading/writing achievement? If we were to develop a new test of reading/writing achievement, we would have to find a valid criterion measure of reading/writing to establish the new test's concurrent validity. Because we know that there are no perfect measures of reading/ writing achievement, we would probably use other achievement tests that are presumed to be valid. Then if our new test elicited test scores correlating significantly with the other tests we would conclude that our new test was a valid measure of achievement. Can we use the correlations found between the teacher assessments and the test scores to establish concurrent validity?

The question may really be, "Do we want to"? First, can we presume the test used by the school district in this study to be a valid one? The technical manual of the test used states that the test is *expected* to correlate significantly with other achievement measures but offers no specific data to support the claim. And how

do we know that the other tests are valid measures? As has already been stated, most reading assessment has not kept pace with advances in reading research, theory, and practice. Even if this particular test correlated highly with other similar tests, would it necessarily be a valid test of reading/ writing as they are conceived of in this school district?

Perhaps a better question might be,"What kinds of information are used in a school district when decisions are being made?" First and second grade teachers who were interviewed about their use of previous end-of-the- year test results and assessments made by their students' previous teachers, unanimously chose the previous teachers' assessments over the test results. Even the principal of the school reported that she rarely used the test results for school-wide instructional decisions. She, too, preferred the more encompassing information obtained from the teachers.

Obviously the results of this study are limited because the population consisted of only one school district. However, having found such consistency of high correlations over the kindergarten, first, and second grade teachers, I believe that the teacher and the test measures can be said to be measuring a number of similar factors. The coefficient of determination for the entire set of classes each year ranged from .45-.50 leading me to believe that the teachers and the test were tapping about fifty percent of the same factors. These relatively high correlations of teacher judgment with standardized tests should ease fears that teacher judgments would be totally at odds with the standardized test results. At the same time, knowing what we do about the negative factors associated with standardized tests and testing in the primary grades and the fact that little use seems to be made of the test results, the data suggest that teacher judgments, based on knowledge of their students' development and knowledge of the processes involved in reading and writing, may be even more valid means of obtaining information for instructional decisions. I urge others to replicate this study. If pupil assessments by teachers in other school districts also correlate highly with test scores, then the notion of subjectivity in the alternative forms may not be a negative factor as some consider it now.

REFERENCES

Bredekamp, S. (Ed.). (1986). *Developmentally appropriate practice.* Washington, D.C.: National Association for the Education of Young Children.

Edelsky, C. & Harman,S. (1988). One more critique of reading tests-with two differences, *English Education* 20, 157-71.

FairTest & NYPIRC. (1990). *Standardized tests and our children: A guide to testing reform in New York.* New York: The National Center for Fair and Open Testing.

Garcia, G. & Pearson, P. D. (1991). The role of assessment in a diverse society. In E. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies.* (pp. 253-278), New York: Teachers College Press.

Harmon, S. (1990). Negative effects of Achievement testing in literacy development. In C. Kamii (Ed.), *Achievement testing in the early grades.* (pp. 111-118), Washington, D.C.: National Association for the Education of Young Children.

Hodges, C. (1989, March). *Readiness tests An obsolete approach to emergent literacy assessment.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hodges, C. (1991). Instruction and assessment of emergent literacy. In L. Weis, P. Altbach, G. Kelly, & H. Petrie (Eds.), *Critical Perspectives on early childhood education* . (pp 153-168), New York: State University of New York Press.

International Reading Association, Early Childhood and Literacy Development Committee. (1986). Literacy development and pre-first grade. *Childhood Education* 63 , 110-11.

Moyer, J. , Egertson, H. & Isenberg, J. (1987). Association for childhood education international position paper: The child-centered kindergarten, *Childhood Education* . 63, 235-42.

National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. *Young Children* 43, 42-47.

Squires, J. (1987, April). Introduction: A special issue on the state of assessment in reading. *The Reading Teacher, 40*, 724-725.

Teale, W. (1989). Developing appropriate assessment of reading and writing in the early childhood classroom. *The Elementary School Journal* 89, 173-183.

Valencia, S. & Pearson P. D. (1986). New models for reading assessment. *Reading Education Report 71.* Urbana, IL: University of Illinois, Center for the Study of Reading.

Table 1. Pearson Correlation Coefficients

CORRELATIONS OF TEACHER ASSESSMENTS WITH TEST SCORES

Kindergarten Teacher	Year One Correlations	First Grade Teacher	Year Two Correlations	Second Grade Teacher	Year Three Correlations
Classroom 1	.62*	Classroom 11	.78*	Classroom 21	.80*
Classroom 2	.86*	Classroom 12	.75*	Classroom 22	.68*
Classroom 3	.72*	Classroom 13	.74*	Classroom 23	.70*
Classroom 4	.87*	Classroom 14	.83*	Classroom 24	.90*
Classroom 5	.85*	Classroom 15	.70*	Classroom 25	.63*
Classroom 6	.72*	Classroom 16	.84*	Classroom 26	.69*
Classroom 7	.58*				
All Kdgs.	.71*	All 1st. Grades	.67*	All 2nd. Grades	.68*

*p<.01

FIGURE ONE

Informal Assessment: Emergent and Early Reading Strategies

NAME: _____

+ = knows * = learning DATES:					
Identifies front of book					
Knows where to start reading					
Aware of page turning direction					
Aware of top-bottom reading					
Aware of left-right					
Aware of return sweep					
Knows punctuation					
period					
question mark					
exclamation mark					
other					
Can identify a letter					
Can identify a word					
Knows print contains message					
Finger pointing					
no attempt					
slides across					
word by word					
Knows Book Terms					
cover					
title					
title page					
author					
illustrator					
page #s					
other					

Notes:

Informal Assessment -Reading Cont'd

NAME:

+ = knows * = learning DATES:					
Story Retelling/Reading					
own version					
retells all important points					
retells parts					
retells almost none					
partially memorized					
memorized					
partially reading print					
reads all print					

Knows g/p correspondence (circle)

b c d f g h j k
l m n p q r s t
v w x y z
a e i o u

Sight words:

FIGURE TWO
Informal Assessment: Writing Strategies

NAME_____

+= knows *= learning DATES:					
Drawing					
simple					
detailed					
dictates story					
Scribble writing					
uncontrolled					
controlled					
left /right					
top/bottom					
random letters					
Copies words					
Invented spelling					
initial consonants					
initial & final					
vowels-incorrect					
vowels-correct					
some standard					
most standard					
Spacing					
strings of letters					
space between letters					
strings of words					
space between words					
Composition					
labels/words					
phrases/sentences					

Informal Assessment: Writing Strategies cont.d.

NAME:

+=knows
*=learning DATES:

Storyline
- new theme each page
- theme continuity
- lit. influence/pattern book
- fiction
- non-fiction

Punctuation
- periods rarely
  - end of sentence
  - overuse
- question mark rarely
  - end of sentence
  - overuse
- exclamation mark rarely
  - end of sentence
  - overuse
- other

Handwriting
- all caps
- caps and lower case

Work Style
- works alone
- asks peers for help
- helps others
- shares and discusses ideas

NOTES:

Informal Assessment: Writing Strategies cont.d.

NAME:_____

+=knows *=learning DATES:					
Interest					
journal writing					
writes on own					
writes when asked					
must be coaxed					
center writing					
chooses to write					
prefers to dictate					
writes books					
Sharing Time					
chooses not to share					
describes picture-simple					
complex					
reads words					
points to words when reading					
uses expression when reading					
adds info when asked					
listens to others					
asks questions of others					
responds with write/book talk					
confused telling and/or asking					

Notes:_____